CAUSAL EXPLANATION AND MANIPULATION

STATHIS PSILLOS

1. INTRODUCTION

Causal explanation proceeds by citing the causes of the *explanandum*. Any model of causal explanation requires a specification of the relation between cause and effect in virtue of which citing the cause explains the effect. In particular, it requires a specification of what it is for the *explanandum* to be causally dependent on the *explanans* and what types of things (broadly understood) the *explanans* are. There have been a number of such models. For the benefit of the unfamiliar reader, here is a brief statement of some major views. On David Lewis's account, *c* causally explains *e* if *c* is connected to *e* with a network of causal chains. For him, causal explanation consists in presenting portions of explanatory information captured by the causal network. On Wesley Salmon's reading, *c* causally explains *e* if *c* is connected with *e* by a suitable continuous causal (i.e., capable of transmitting a mark) process. On the standard deductive-nomological reading of causal explanation, for *c* to causally explain *e*, *c* must be a nomologically sufficient condition for *e*. And for John Mackie, for *c* to causally explain *e* there must be event-types *C* and *E* such that *C* is an inus-condition for *E*.[53]

In a series of papers and a book, James Woodward (1997, 2000, 2002, 2003a, 2003b) has put forward a 'manipulationist' account of causal explanation. Briefly put, *c* causally explains *e* if *e* causally depends on *c*, where the notion of causal dependence is understood in terms of relevant (interventionist) counterfactual, that is counterfactuals that describe the outcomes of interventions. A bit more accurately, *c* causally explains *e* if, were *c* to be (actually or counterfactually) manipulated, *e* would change too. This model ties causal explanation to actual and counterfactual experiments that show how manipulation of factors mentioned in the *explanans* would alter the *explanandum*. It also stresses the role of invariant relationships, as opposed to strict laws, in causal explanation. Explanation in this model consists in answering a network of "what-if-things-had-been-different questions", thereby placing the *explanandum* within a pattern of counterfactual dependencies (cf. Woodward 2003a, p. 201). For instance, the law of ideal gases is said to be explanatory not because it renders a certain *explanandum* (e.g., that the pressure of a certain gas increased) nomically expected, but because it can tell us how the

[53] For details on all these, see my (2002)

J. Persson and P. Ylikoski (eds.), Rethinking Explanation, 93–107.

pressure of the gas would have changed, had the antecedent conditions (e.g., the volume of the gas) been different. The explanation proceeds by locating the *explanandum* "within a space of alternative possibilities" (Woodward 2003a, p. 191). The key idea, I take it, is that causal explanation shows how the *explanandum* depends on the *explanans* in stable way. Not only does it show why the *explanandum* holds; it also shows how it would *vary* in a stable way, had the factors mentioned in the *explanans* been different.

Woodward's theory is developed in great detail in his (2003a) and I cannot do full justice to it in this paper. Since I will be mostly critical of his appeal to interventionist counterfactual conditionals, I should state right from the start that Woodward's theory is invariably interesting and insightful. In particular, it casts new light on the practice of causal explanation, especially in the so-called special sciences. It makes clear how causal explanation is concerned with factors that make a difference to the presence or absence of the *explanandum*. It deals quite effectively with the traditional problems of asymmetries in explanation and of citing irrelevant factors as explanatory. It accommodates omissions and preventings as explanatory factors. It shows that not all causal explanations should take the form of deductive (or inductive) arguments. Nonetheless, it displays how generalisations (suitably understood) do play a role in causal explanation.

Leaving all these positive elements to one side, this paper will focus on two central conceptual ingredients of Woodward's account of causal explanation, viz., *interventionist* counterfactuals and *invariant* generalisations. Section 2 offers a brief presentation of Woodward's theory highlighting its two central ingredients. Section 3 calls into question Woodward's interventionist counterfactuals. It claims that they blur the distinction between truth- and evidence-conditions of counterfactual assertions and leave us with no clear account of the semantics of counterfactuals. Sections 4 discusses the role of laws in causal explanation and claims that the very possibility of experimental counterfactuals requires that laws are understood in a sense stronger than relations of invariance among variables.

2. MANIPULATIONIST CAUSAL EXPLANATION

Woodward takes his theory of causal explanation to be intimately linked to his theory of causation. This, of course, is as it should be. Causal explanation is meant to provide information about the *causes* of the *explananda*, and hence it requires an account of what it is for *c* to cause *e*. On Woodward's view, causation is based on counterfactual manipulation. His theory is *counterfactual* in the following sense: what matters is what *would* happen to a relationship, *were* interventions to be carried out. A relationship among some variables *X* and *Y* is causal if, were there an intervention that changed the value of *X* appropriately, the relationship between *X* and *Y* wouldn't change *and* the value of *Y* would change. To use a stock example, the force exerted on a spring *causes* a change of its length, because if an intervention changed the force exerted on the spring, the length of the spring would change too (but the relationship between the two magnitudes—expressed by Hooke's law—would remain invariant, within a certain range of interventions).

Let us describe, somewhat sketchily, the two key notions of intervention and invariance. The gist of Woodward's characterisation of an *intervention* is this. A change of the value of X counts as an intervention I if it has the following characteristics:

a) the change of the value of X is entirely due to the intervention I;
b) the intervention changes the value of Y, if at all, only through changing the value of X.

The *first* characteristic makes sure that the change of X does not have causes other than the intervention I, while the *second* makes sure that the change of Y does not have causes other than the change of X (and its possible effects).[54] These characteristics are meant to ensure that Y-changes are exclusively due to X-changes, which, in turn, are exclusively due to the intervention I. As Woodward stresses, there is a close link between intervention and manipulation. Yet, his account makes no special reference to human beings and their (manipulative) activities. In so far as a process has the right characteristics, it counts as an intervention. So interventions can occur 'naturally', even if they can be highlighted by reference to "an idealised experimental manipulation" (2000, p. 199).

Woodward links the notion of intervention with the notion of *invariance*. A certain relation (or a generalisation) is invariant, Woodward says, "if it would continue to hold—would remain stable or unchanged—as various other conditions change" (2000, p. 205). What really matters for the characterisation of invariance is that the generalisation remains stable under a set of actual and counterfactual *interventions*. So Woodward (2000, p. 235) notes:

> the notion of invariance is obviously a modal or counterfactual notion [since it has to do] with whether a relationship would remain stable if, perhaps contrary to actual fact, certain changes or interventions were to occur.

Let me highlight three important general elements of Woodward's approach. *First*, causal claims relate variables. He (2003a, p. 112) insists that causes should be such that it makes sense to say of them that they could be changed or manipulated. Thinking of them as variables, which can take different values, is then quite natural. But as he goes on to note, it is not difficult to translate talk in terms of changes in the values of variables into talk in terms of events and conversely. For instance, instead of saying that the hitting by the hammer (an event) caused the shattering of the vase (another event), we may say that the change of the value of a certain indicator variable from *not-hit* to *hit* caused the change of the value of another variable from *unshattered* to *shattered*. This strategy, however, will not work in cases in which putative causes cannot be understood as values of variables.[55] But then again, this is

[54] There is a *third* characteristic too, viz., that the intervention I is not correlated with other causes of Y besides X.

[55] For an important attempt to show how the relata of the interventionist counterfactual approach can be seen as events, see Kluve (2004, especially 81-2).

fine for Woodward, as he claims that in those cases causal claims will be, to say the least, ambiguous (cf. 2003a, p. 115ff).

Second, generalisations need not be invariant under *all* possible interventions. Hooke's law, for instance, would 'break down' if one intervened to stretch the spring beyond its breaking point. Still, Hooke's law does remain invariant under some set of interventions. In so far as a generalisation is invariant under a certain range of interventions, it can be explanatorily useful, without being exceptionless (cf. 2000, p. 227-8). Woodward (2000, p. 214) stresses: "[t]here are generalisations that are invariant and that can be used to answer a range of what-if-things-had-been-different questions and that hence are explanatory, even though we may not wish to regard them as laws and even though they lack many of the features traditionally assigned to laws by philosophers". In particular, a generalisation can be *causal* even if it is not universally invariant (cf. 2003a, p. 15).

Third, Woodward does not aim to offer a reductive account of causation or causal explanation. The notion of intervention is itself causal and, in any case, causal considerations are necessary to specify when a relationship among some variables is causal. For instance, an appropriate intervention I on variable X with respect to variable Y should be such that it is not correlated with other *causes* of Y or does not directly *cause* a change of the value of Y. I think Woodward (2003a, p. 104-7) is right in insisting that his account is not trapped in a vicious circle. In any case, an account of causation or causal explanation need not be reductive to be illuminating.

In light of the above, causal explanation proceeds by exploiting the manipulationist element of causation and the invariant element of generalisations. Explanatory information "is information that is potentially relevant to manipulation and control" (Woodward 2003a, p. 10). Causal relations are explanatory because they provide information about counterfactual dependencies among causal variables. And invariant generalisations are explanatory because they exhibit stable patterns of counterfactual dependence among causal variables in virtue of which different values of the effect-variable counterfactually depend on different values of the cause-variable.

3. COUNTERFACTUALS

It is already evident that counterfactual conditionals loom large in Woodward's account. Interventions need not be actual. They can be hypothetical or counterfactual. And invariance is not understood in terms of stability under actual interventions. The causal relationship (generalisation) should be invariant under hypothetical or counterfactual interventions.

Counterfactual conditionals have been reprimanded on the ground that they are context-dependent and vague. Take, for instance, the following counterfactual: 'If Jones had not smoked so heavily, he would have lived a few years more'. What is it for it to be true? Any attempt to say whether it is true, were it to be possible at all, would require specifying what else should be held fixed. For instance, other aspects of Jones's health should be held fixed, assuming that other factors (e.g., a weak heart) wouldn't cause a premature death, anyway. But what things to hold fix is not,

necessarily, an objective matter. Or, consider the following pair of counterfactuals: 'If Julius Caesar had been in charge of U. N. Forces during the Korean war, then he would have used nuclear weapons' and 'If Julius Caesar had been in charge of U. N. Forces during the Korean war, then he would have used catapults'. It is difficult to see how we could possibly tell which of them, if any, is true.

As the reader will surely know, there have been many significant attempts to offer semantic for counterfactual conditionals. Perhaps the most well-developed, and certainly the most well-known, is Lewis's (1973) account in terms of possible worlds. I will not discuss this theory here.[56] The relevant point is that Woodward offers an account of counterfactuals that tries to avoid the metaphysical excesses of Lewis's theory.[57]

3.1 Experimental counterfactuals

Woodward is very careful in his use of counterfactuals. Not all of them are of the right sort for the evaluation of whether a relation is causal. Only counterfactuals that are related to *interventions* can be of help. An intervention gives rise to an "active counterfactual", that is, to a counterfactual whose antecedent is made true "by interventions" (1997, p. 31; 2000, p. 199). In his (2003a, p. 122) he stresses that

> the appropriate counterfactuals for elucidating causal claims are not just any counterfactuals but rather counterfactuals of a very special sort: those that have to do with the outcomes of hypothetical interventions. […] it does seem plausible that counterfactuals that we do not know how to interpret as (or associate with) claims about the outcomes of well-defined interventions will often lack a clear meaning or truth value.

In his (2003b, p. 3), he very explicitly characterises the appropriate counterfactuals in terms of *experiments*: they "are understood as claims about what would happen if a certain sort of experiment were to be performed" (cf. also 2003a, p. 10 and 114).

Consider a case he (2003b, p. 4-5) discusses. Take Ohm's law (that the voltage E of a current is equal to the product of its intensity I times the resistance R of the wire) and consider the following two counterfactuals:

(1) If the resistance were set to $R=r$ at time t, and the voltage were set to $E=e$ at t, then the intensity I would be $i=e/r$ at t.
(2) If the resistance were set to $R=r$ at time t, and the voltage were set to $E=e$ at time t, then the intensity I would be $i^* \neq e/r$ at t.

There is nothing mysterious here, says Woodward, "as long as we can describe how to test them" (2003b, p. 6). We can perform the experiments at a future time t^* in order to see whether (1) or (2) is true. If, on the other hand, we are interested in what *would* have happened had we performed the experiment in a past time t, Woodward invites us to rely on the "very good evidence" we have "that the

[56] See my (2002, 92-101).
[57] For a discussion of Lewis's theory in relation to Woodward's see his (2003a, 133-45).

behaviour of the circuit is stable over time" (2003b, p. 5). Given this evidence, we can assume, in effect, that the *actual* performance of the experiment at a future time *t** is as good for the assessment of (1) and (2) as a *hypothetical* performance of the experiment at the past time *t*.

For Woodward, the truth-conditions of counterfactual statements (and their truth-values) are not specified by means of an abstract metaphysical theory, e.g., by means of abstract relations of similarity among possible worlds. He calls his own approach "pragmatic". That's how he (2003b, p. 4) puts it:

> For it to be legitimate to use counterfactuals for these goals [understanding causal claims and problems of causal inference], I think that it is enough that (a) they be useful in solving problems, clarifying concepts, and facilitating inference, that (b) we be able to explain how the kinds of counterfactual claims we are using can be tested or how empirical evidence can be brought to bear on them, and (c) we have some system for representing counterfactual claims that allows us to reason with them and draw inferences in a way that is precise, truth-preserving and so on.

Yet, Woodward's view is also meant to be realist and objectivist. He is quite clear that counterfactual conditionals have non-trivial truth-values independently of the actual and hypothetical experiments by virtue of which it can be assessed whether they are true or false. He (2003b, p. 5) says:

> On the face of things, doing the experiment corresponding to the antecedent of (1) and (2) doesn't *make* (1) and (2) have the truth values they do. Instead the experiments look like ways of *finding out* what the truth values of (1) and (2) were all along. On this view of the matter, (1) and (2) have non-trivial truth values—one is true and the other false—even if we don't do the experiments of realizing their antecedents. Of course, we may not *know* which of (1) and (2) is true and which false if we don't do these experiments and don't have evidence from some other source, but this does not mean that (1) and (2) both have the same truth-value.

This point is repeated in his (2003a, p. 123), where he stresses:

> We think instead of [a counterfactual such as (1) above] as having a determinate meaning and truth value whether or not the experiment is actually carried out—it is precisely because the experimenters want to *discover* whether [this counterfactual] is true or false that they conduct the experiment.

So though "pragmatic", Woodward's theory is also objectivist. But it is minimally so. As he (2003a, p. 121-2) notes, his view:

> requires only that there be facts of the matter, independent of facts about human abilities and psychology, about which counterfactual claims about the outcome of hypothetical experiments are true or false and about whether a correlation between *C* and *E* reflects a causal relationship between *C* and *E* or not. Beyond this, it commits us to no particular metaphysical picture of the 'truth-makers' for causal claims.

The main problem that I see in Woodward's theory relates to the question: *what are the truth-conditions of counterfactual assertions?* Woodward doesn't take all counterfactuals to be meaningful and truth-valuable. As we have seen (see also 2003a, 122), he takes only a subclass of them, the active counterfactuals, to be such. However, he does not want to say that the truth-conditions of active counterfactuals are fully specified by (are reduced to) actual and hypothetical experiments. If he said this, he could no longer say that active counterfactuals have

determinate truth-conditions *independently* of the (actual and hypothetical) experiments that can test them. In other words, Woodward wants to distinguish between the truth-conditions of counterfactuals and their evidence-(or test) conditions, which are captured by certain actual and hypothetical experiments. The problem that arises is this. Though we are given a relatively detailed account of the evidence-conditions of counterfactuals, we are not given anything remotely like this for their *truth-conditions*. What, in other words, is it that makes a certain counterfactual conditional true?

A thought here might be that there is no need to say anything more detailed about the truth-conditions of counterfactuals than offering a Tarski-style meta-linguistic account of them of the form

(*T*)
'If *x* had been the case, then *y* would have been the case' is true iff if *x* had been the case, then *y* would have been the case.

This move is possible but not terribly informative. We don't know when to assert (or hold true) the right hand-side. And the question is precisely this: when is it right to assert (or hold true) the right-hand side? Suppose we were to tell a story in terms of actual and hypothetical experiments that realise the antecedent of the right-hand side of (*T*). The problem with this move is that the truth-conditions of the counterfactual conditional would be specified in terms of its evidence-conditions, which is exactly what Woodward wants to block. Besides, if we just stayed with (*T*) above, without any further explication of its right-hand side, *any* counterfactual assertion (and not just the active counterfactuals) would end up meaningful and truth-valuable. Here again, Woodward's project would be undermined. Woodward is adamant: "Just as non counterfactual claims (e.g., about the past, the future, or unobservables) about which we have no evidence can nonetheless possess non-trivial truth-values, so also for counterfactuals" (2003b, p. 5). This is fine. But in the case of claims about the past or about unobservables there are well-known stories to be told as to what the difference is between truth- and evidence-conditions. When it comes to Woodward's counterfactuals, we are *not* told such a story.

Another thought might be motivated by Woodward's view that causal claims are irreducible. Woodward says:

> According to the manipulationist account, given that *C* causes *E*, which counterfactual claims involving *C* and *E* are true will always depend on which other *causal* claims involving other variables besides *C* and *E* are true in the situation under discussion. For example, it will depend on whether other causes of *E* besides *C* are present (2003a, p. 136).

The idea here, I take it, is that the truth-conditions of counterfactuals depend on the truth-conditions of certain causal claims, most typically causal claims about the larger causal structure in which the variables that appear in the counterfactuals under examination are embedded. Intuitively, this is a cogent claim. Consider two variables *X* and *Y* and examine the counterfactual: if *X* had changed (that is, if an intervention *I* had changed the value of *X*), the value of *Y* would have changed. Whether this is true or false will depend on whether *I* causes the value of *Y* to

change by a route independent of X, or on whether some other variable Z causes a direct change of the value of Y. Causal facts such as these are part of the truth-conditions of the foregoing counterfactual. It is clear that they may, or may not, obtain independently of any intervention on X. So whether on not an intervention I on X were to occur, it might be the case that were it to occur, it would not influence the value of Y by a route independent of X. The thought, then, may be that the truth-conditions of a counterfactual are specified by certain causal facts that involve the variables that appear in the counterfactual as well as the variables of the broader causal structure in which the variables of interest are embedded.

I see two problems with this thought. The *first* is that this account is very abstract and general. It *is* informative since it says that causal facts are required for the truth of counterfactuals, but what these facts are will depend on, and vary with, each causal structure under consideration. So the proposal does not specify which causal facts are required for the truth of counterfactuals. What these facts are will depend on each particular causal structure.

The *second* problem is that this account seems circular. Causal claims, we are told, should be understood in terms of counterfactual dependence (where the counterfactuals are interventionist). To fix our ideas, let us consider the causal claim

B_0: X causes Y.

For B_0 to be true, the following counterfactual C_1 should be true.

C_1: if X had changed (that is, if an intervention I had changed the value of X), the value of Y would have changed.

On the thought we are presently considering, the truth of C_1 will depend, among other things, on the truth of another causal claim:

B_1: I does not cause a change to the value of Y directly, (that is, by a route independent of X).

How does the truth of B_1 depend on counterfactuals? Let us assume that relations of counterfactual dependence are *part* of the truth-conditions of causal claims. Then, at least *another* (interventionist) counterfactual C_2 would have to be true in order for B_1 to be true.

C_2: if an(other) intervention I' had changed the value of I, the value of Y would not have changed (by a route independent of X).

But what makes C_2 true? Suppose it is another causal claim B_2.

B_2: I' does not cause a change to the value of Y directly.

For B_2 to be true, another counterfactual C_3 would have to be true, and so on. Either a regress is in the offing or the truth of some causal claims has to be accepted as a brute fact. In the former case, counterfactuals are part of the truth-conditions of other counterfactuals, with no independent account of what it is for a counterfactual to be true. In the latter case, we are left in the dark as to what causal claims capture brute facts. In particular, why should we not take it as a brute fact that B_0 or B_1 is true?

Suppose, on the other hand, that we do *not* take relations of counterfactual dependence to be part of the truth-conditions of causal claims. We would still need an account of the truth-conditions of causal claims. But even if we ignore this, a circle is still present. Suppose we settle for the weaker view that relations of counterfactual dependence are needed for establishing that a causal claim is true (without being them that *make* this claim true). The circle we are now caught in is this: establishing that certain counterfactuals are true is necessary for establishing that other counterfactuals are true or false. For instance, for establishing the claim that C_1 is true, it is required that another counterfactual C_2 is established as true and so on. Since C_1 is distinct from C_2, the circle *might* not be vicious. But the point is that there is no obvious place to break the circle of counterfactuals and make it going.

We have examined two ways to specify the truth-conditions for counterfactual claims and we have found them both wanting. Still, there are two general options available. One is to *collapse* the truth-conditions of counterfactuals to their evidence-conditions. One can see the *prima facie* attraction of this move. Since evidence-conditions are specified in terms of actual and hypothetical experiments, the right sort of counterfactuals (the active counterfactuals) *and only those* end up being meaningful and truth-valuable. But there is an important drawback. Recall counterfactual assertion (1) above. On the option presently considered, what makes (1) true is that its evidence-conditions obtain. Under this option, counterfactual conditionals lose, so to speak, their counterfactuality. (1) becomes a shorthand for a future prediction and/or the evidence that supports the relevant law. If *t* is a *future* time, (1) gives way to an actual conditional (a prediction). If *t* is a past time, then, given that there is good evidence for Ohm's law, all that (1) asserts under the present option is that there has been good evidence for the law.

In any case, Woodward is keen to keep evidence- and truth-conditions apart. Then, (and this is the other option available) some informative story should be told as to what the truth-conditions of counterfactual conditionals *are* and *how* they are connected with their evidence-conditions (that is, with actual and hypothetical experiments). There may be a number of stories to be told here.[58] The one I favour

[58] One might try to keep truth- and evidence-conditions apart by saying that counterfactual assertions have excess content over their evidence-conditions in the way in which statements about the past have excess content over their (present) evidence-conditions. Take the view (roughly Dummett's) that statements about the past are meaningful and true in so far as they are verifiable (i.e., their truth can be known). This view may legitimately distinguish between the *content* of a statement about the past and the present or future evidence there is for it. Plausibly, this excess content of a past statement may be cast in terms of counterfactuals: a meaningful past statement *p* implies counterfactuals of the form 'if *x* were present at time *t*, *x* would verify that *p*'. This move presupposes that there are meaningful and

ties the truth-conditions of counterfactual assertions to *laws of nature*. It is then easy to see how the evidence-conditions (that is, actual and hypothetical experiments) are connected with the truth-conditions of a counterfactual: actual and hypothetical experiments are symptoms for the presence of a law. There is a hurdle to be jumped, however. It is notorious that many attempts to distinguish between genuine laws of nature and accidentally true generalisations rely on the claim that laws do, while accidents do not, support counterfactuals. So counterfactuals are called for to distinguish laws from accidents. If at the same time laws are called for to tell when a counterfactual is true, we go around in circles. Fortunately, there is the Mill-Ramsey-Lewis view of laws (see my 2002, Chapter 5). Laws are those regularities which are members of a coherent system of regularities, in particular, a system which can be represented as an ideal deductive axiomatic system striking a good balance between *simplicity* and *strength*. On this view, laws are identified independently of their ability to support counterfactuals. Hence, they can be used to specify the conditions under which a counterfactual is true.[59]

It might be that Woodward aims only to provide a *criterion* of meaningfulness for counterfactual conditionals without also specifying their truth-conditions. This would seem in order with his "pragmatic" account of counterfactuals, since it would offer a criterion of meaningfulness and a description of the 'evidence conditions' of counterfactuals, which are presumed to be enough to understand causation and causal explanation. In response to this, I would not deny that Woodward has indeed offered a sufficient condition of meaningfulness. Saying that counterfactuals are meaningful if they can be interpreted as claims about actual and hypothetical experiments is fine. But can this also be taken as a necessary condition? Can we say that *only* those counterfactuals are meaningful which can be seen as claims for actual and hypothetical experiments? If we did say this, we would rule out as meaningless a number of counterfactuals that philosophers have played with over the years, e.g., the pair of Julius Caesar counterfactuals considered in section 3. Though I agree with him that they are "unclear", I am not sure they are meaningless. Take one of Lewis's examples, that had he walked on water, he would not have been wet. I don't think it is meaningless. One may well wonder what the point of offering such counterfactuals might be. But whatever it is, they are understood and, perhaps, are true. Perhaps, as Woodward (2003a, p. 151) says, the antecedents of such counterfactuals are "unmanipulable for conceptual reasons". But if they are understood (and if they are true), this would be enough of an argument *against* the view that manipulability offers a necessary condition for meaningfulness.

It turns out, however, that there are more sensible counterfactuals that fail Woodward's criterion. Some of them are discussed by Woodward himself (2003a, p. 127-33). Consider the true causal claim: Changes in the position of the moon with

---

true counterfactual assertions. But note that a similar story *cannot* be told about counterfactual conditionals. If we were to treat their supposed excess content in the way we just treated the excess content of past statements, we would be involved in an obvious regress: we would need counterfactuals to account for the excess content of counterfactuals.

[59] Obviously, the same holds for the Armstrong-Dretske-Tooley view of laws (see my 2002, chapter 6). If one takes laws as necessitating relations among properties, then one can explain why laws support counterfactuals and, at the same time, identify laws *independently* of this support.

respect to the earth and corresponding changes in the gravitational attraction exerted by the moon on the earth's surface cause changes in the motion of the tides. As Woodward adamantly admits, this claim cannot be said to be true on the basis of interventionist (experimental) counterfactuals, simply because realising the antecedent of the relevant counterfactual is physically impossible. His response to this is an alternative way for assessing counterfactuals. This is that counterfactuals can be meaningful if there is some "basis for assessing the truth of counterfactual claims concerning what would happen if various interventions were to occur". Then, he adds, "it doesn't matter that it may not be physically possible for those interventions to occur" (2003a, p. 130). And he sums it up by saying that "an intervention on *X* with respect to *Y* will be 'possible' as long it is logically or conceptually possible for a process meeting the conditions for an intervention on *X* with respect to *Y* to occur" (2003a, p. 132). My worry then is this. We now have a much more liberal criterion of meaningfulness at play, and it is not clear, to say the least, which counterfactuals end up meaningless by applying it.

In any case, Woodward (2003a, p. 132) offers an important warning:

> [I]t would be a mistake to make the physical possibility of an intervention on *C* constitutive in any way of what it is for there to be a causal connection between *C* and *E*. […] When an intervention changes *C* and in this way changes *E*, this exploits an independently existing causal link between *C* and *E*. One can perfectly well have the link without the physical possibility of an intervention on *C*.

I take this to imply that his counterfactual approach provides an *extrinsic* way to identify a sequence of events as causal, viz., that the sequence remains invariant under certain interventions. In an earlier piece, he (2000, p. 204) stressed:

> what matters for whether *X* causes […] *Y* is the 'intrinsic' character of the *X*-*Y* relationship but the attractiveness of an intervention is precisely that it provides an extrinsic way of picking out or specifying this intrinsic feature.

So there seems to be a conceptual distinction between causation and invariance-under-interventions: there is an *intrinsic* feature of a relationship in virtue of which it is causal, an *extrinsic* symptom of which is its invariance under interventions.[60] If I have got Woodward right, causation has excess content over invariance-under-interventions. So there is more to causation—*qua* an intrinsic relation—than invariance-under-actual-and-counterfactual-interventions. Hence, there is more to be understood about what causation and causal explanation are.

To sum up. We need to be told more about the truth-conditions of counterfactual conditionals. If Woodward ties too close a knot between counterfactuals and actual and hypothetical experiments, then counterfactual assertions may reduce to claims about actual and hypothetical experiments (without any excess content). If, on the other hand, Woodward wants to insist that counterfactuals have their truth-conditions independently of their evidence-conditions, then it is an entirely open option that the truth-conditions of counterfactual assertions involve laws of nature.

[60] In his (2003a, p. 125) Woodward says "there is a certain kind of relationship with intrinsic features that we exploit or make use of when we bring about *B* by bringing about *A*".

3.2 No laws in, no counterfactuals out

As we have already seen, when it comes to causal explanation, Woodward stresses that reliance on invariant generalisations is enough for it. He (2003a, p. 236) says:

> [W]hat matters for whether a generalisation is explanatory is whether it can be used to answer a range of what-if-things-had-been-different questions and to support the right sorts of counterfactuals about what will happen under interventions".

Naturally, when checking whether a generalisation or a relationship among magnitudes or variables is invariant we need to subject it to some variations/changes/interventions. What changes will it be subjected to? The obvious answer is: those that are permitted, or are permissible, by the laws of nature. Suppose that we test Ohm's law. Suppose also that one of the interventions envisaged was to see whether it would remain invariant, if the measurement of the intensity of the current was made on a spaceship, which moved faster than light. This, of course, cannot be done, because it is a *law* that nothing travels faster than light. So, some *laws* must be in place before, based on considerations of invariance, it is established that some generalisation is invariant under some interventions. Hence, Woodward's notion of "invariance under interventions" cannot offer an adequate analysis of lawhood, since laws are required to determine what interventions are possible.

Couldn't Woodward say that even basic laws—those that determine what interventions and changes are possible—express just relations of invariance? Take, once more, the law that nothing travels faster than light. Can the fact that it is a law be the result of subjecting it to interventions and changes? Hardly. For it itself establishes the *limits* of possible interventions and control.[61] I do not doubt that it may well be the case that genuine laws express relations of invariance. But this is not the issue. For, the manifestation of invariance might well be the *symptom* of a law, without being constitutive of it.

It seems that Woodward must be committed to this symptom/constitution distinction. As he explains in detail, invariance does not characterise laws only; other relationships or generalisations, which cannot be deemed laws, display invariance, especially in the special sciences. For instance, Woodward (2000, p. 214) notes:

> [t]here are generalisations that are invariant and that can be used to answer a range of what-if-things-had-been-different questions and that hence are explanatory, even though we may not wish to regard them as laws and even though they lack many of the features traditionally assigned to laws by philosophers.

Note, however, that at least some accidental generalisations do possess *some* range of invariance. So if invariance is to be found in laws as well as in non-laws, it should be at best a *symptom* of lawhood. What, then, does lawhood consist in? Woodward is perfectly happy with the thought that laws are not what philosophers have taken them to be. He (2000, p. 222) thinks that most of the standard criteria

[61] Woodward (2000, p. 206-7) too agrees that this law cannot be accounted for in terms of invariance.

> are not helpful either for understanding what is distinctive about laws of nature or for understanding the feature that characterise explanatory generalisations in the special sciences.

In particular, he takes it that in so far as a generalisation is invariant under a certain range of interventions, it can be a law without being exceptionless (cf. 2000, p. 227-8). But no clear picture emerges as to what exactly makes a generalisation a law. For, as Woodward (2000, p. 227) admits, even laws will *not* be invariant under *all* actual and possible interventions. For instance, Maxwell's laws break down at the Planck scale, where quantum mechanical effects take over. As a result of all this, the difference between laws, invariant-generalisations-that-are-explanatorily-useful-but-non-laws, and mere accidents is deemed to be a difference "in degree (…) rather than of kind" (2000, p. 241). It is a difference in degree precisely because the notion of invariance under interventions admits of degrees. Some generalisations have a wider range of invariance, whereas others have a narrower range and yet others are "highly non-invariant" (2000, p. 237). This is not to say, Woodward claims, that the difference in degree is no difference at all. For, as he (2000, p. 242) says,

> the features possessed by generalisations, like Maxwell's equations [which are paradigmatic cases of laws]—greater scope and invariance under larger, more clearly defined, and important classes of interventions and changes—represent just the sort of generality and unconditionality standardly associated with laws of nature.

Be that as it may, it should be stressed that laws are required in order to fix the range of invariance of a generalisation. For, in order to specify the range of invariance of a generalisation, we first need a) to specify what interventions are physically possible and b) which of them, if they happened, would leave the given generalisation unchanged. Both of the above, however, need a prior reliance on *laws*. As noted above, it is laws that specify the physically possible interventions. What needs to be added here is that it is laws that govern the assessment of the counterfactual in (b). For instance, specifying what interventions, had they happened, would have left Kepler's law unchanged requires holding other *laws* fixed. For if laws, e.g., Newton's laws, were allowed to be violated, then the range of invariance of Kepler's laws would be very limited. So, it seems that Woodward's account boils down to the following circular statement: a generalisation is a *law* if it is invariant "under a large and important set of changes" (2000, p. 241), where the relevant set of changes is determined by *laws*.[62]

To sum up. Without an independent account of what laws are, there is no clear way in which we can deem some (interventionist) counterfactual assertions true or false. Which interventions are physically possible and which interventions leave certain relations invariant depends on what laws there are. The latter cannot be fully understood as relations that remain invariant under interventions since they specify what interventions are possible.

[62] I take to heart Marc Lange's (2000) recent important diagnosis: either *all* laws, taken as a whole, form an invariant-under-interventions set, or, strictly speaking, no law, taken in isolation, is invariant-under-interventions. This does not yet tell us what laws *are*. But it does tell us what marks them off from intuitively accidental generalisations.

4. CONCLUSION

Perhaps, the worries raised in this paper do not affect causal explanation as a practical activity. In many practical cases, we may well have a lot of information about a particular causal structure and this may be enough to answer questions about which (interventionist) counterfactuals are true and what generalisations are invariant under interventions. When we deal with *stable causal or nomological structures*[63] interventionist counterfactuals are meaningful and truth-valuable. The worries raised in the paper concern the prospects of the manipulationist account as a philosophical theory of causal explanation. Simply put, the main worry is that, as it stands, Woodward's theory highlights and exploits the *symptoms* of a good causal explanation, without offering a fully-fledged theory of what causal explanation consists in. Invariance-under-interventions is a symptom of causal relations and laws. It is not what causation or lawhood consists in. It is a great virtue of Woodward's approach that exploits these symptoms to show how causal explanation can proceed. But this undeniable virtue should not obscure the fact that there is more to causal explanation (by there being more to causation and to lawhood) than stable relations of (interventionist) counterfactual dependence.

Woodward (2003a, p. 114 and 130) has stressed that his notion of intervention should be seen as a "regulative ideal". Its function, he says, is "to characterise the notion of an ideal experimental manipulation and in this way to give a purchase on what we mean or are trying to establish when we claim that *X* causes *Y*" (2003a, p. 130). Perhaps, his theory of causal explanation is also meant to be regulative ideal: it tells us what we should mean and strive to do when we claim that *X* causally explains *Y*. I have no quarrel with this, provided it is also acknowledged that the regulative ideal is still short of being constitutive of what causal explanation is.

REFERENCES

Kluve, J. (2004). On the Role of Counterfactuals in Inferring Causal Effects. *Foundations of Science* 9: 65-101.

Lange, M. (2000). *Natural Laws in Scientific Practice*. Oxford: Oxford University Press.

Lewis, D. (1973). *Counterfactuals*. Cambridge MA: Harvard University Press.

Psillos, S. (2002). *Causation and Explanation*. Chesham: Acumen.

Simon, H. A. and Rescher, N. (1966). Cause and Counterfactual. *Philosophy of Science* 33: 323-40.

Woodward, J. (1997). Explanation, Invariance and Intervention. *Philosophy of Science* 64 (Proceedings): 26-41.

Woodward, J. (2000). Explanation and Invariance in the Special Sciences. *The British Journal for the Philosophy of Science* 51: 197-254.

[63] My favourite way to spell out this notion is given by Simon and Rescher (1966). In fact, in showing how a stable structure can make some counterfactuals true, they blend the causal and the nomological in a fine way.

Woodward, J. (2002). What is a Mechanism? A Counterfactual Account. *Philosophy of Science* 69: 366-377.
Woodward, J. (2003a). *Making Things Happen: A Theory of Causal Explanation*.
New York: Oxford University Press.
Woodward, J. (2003b). Counterfactuals and Causal Explanation. http://philsciarchive.pitt.edu/archive/00000839/.